Wierstorf, Hagen; Raake, Alexander; Spors, Sascha:

Assessing localization accuracy in sound field synthesis

Original published in:

The journal of the Acoustical Society of America : JASA-O. - Melville, NY : AIP Publ. - 141 (2017), 2, p. 1111-1119.

ISSN: 1520-8524

DOI: 10.1121/1.4976061

URL: https://doi.org/10.1121/1.4976061

[*Visited:* 2017-04-13]

Assessing localization accuracy in sound field synthesis[a)]

Hagen Wierstorf[b),c)] and Alexander Raake[c)]
Assessment of IP-based Applications, Technische Universität Berlin, Berlin DE-10587, Germany

Sascha Spors
Institute of Communications Engineering, University of Rostock, Rostock DE-18119, Germany

(Received 5 July 2016; revised 15 January 2017; accepted 20 January 2017; published online 27 February 2017)

Sound field synthesis methods like Wave Field Synthesis (WFS) and Near-Field Compensated Higher Order Ambisonics synthesize a sound field in an extended area surrounded by loudspeakers. Because of the limited number of applicable loudspeakers the synthesized sound field includes artifacts. This paper investigates the influence of these artifacts on the accuracy with which a listener can localize a synthesized source. This was performed with listening tests using dynamic binaural synthesis to simulate different sound field synthesis methods and incorporated several listening positions. The results show that WFS is able to provide good localization accuracy in the whole listening area even for a low number of loudspeakers. For Near-Field Compensated Higher Order Ambisonics the achievable localization accuracy of the listener depends highly on the Ambisonics order and shows large localization deviations for low orders, where splitting of the perceived sound source was sometimes reported.

[http://dx.doi.org/10.1121/1.4976061]

[GCS] Pages: 1111–1119

I. INTRODUCTION

The goal of sound field synthesis techniques is to create a predefined sound pressure in an extended area that is surrounded by loudspeakers.[1,2] Depending on the distances between the loudspeakers and the applied techniques, the synthesized sound field deviates from the desired one.[3,4] This paper investigates the influence of those deviations on the ability of listeners to localise sound sources in the synthesized sound field. The listeners judged the perceived horizontal direction for different sound field synthesis techniques, loudspeaker distances, synthesized sound source types, and listening positions. The results identify the needed physical parameters of a sound field synthesis system to achieve high localization accuracy.

The human auditory system has the remarkable ability to discriminate changes in the horizontal direction of a sound source as small as 1°.[5] This imposes strict requirements on a sound field synthesis system, if the system tries to enable listeners to discriminate synthesized sources with the same accuracy. Several studies investigated localization accuracy for different sound field synthesis setups, but in most of them only a central listening position was considered. No studies performed a systematic comparison between different sound field synthesis techniques in the whole listening area.

For the well-established sound field synthesis technique of Wave Field Synthesis (WFS), results show that localization is at most slightly impaired for loudspeaker spacings less than 25 cm. Those results were obtained for different linear loudspeaker arrays synthesizing a point source[3,6,7] or a focused source, which is a source in front of the loudspeaker array.[8] The listeners were always placed at a central listening position. In a recent publication, Wierstorf *et al.*[9] have investigated the localization at 16 different listening positions for a linear loudspeaker array. The results demonstrate the possibility of WFS to ensure similar localisation performance in the whole listening area. For a loudspeaker spacing of 20 cm, no difference to the localization of a real source was found.

Another sound field synthesis technique, which is only available for circular or spherical geometries is Near-Field Compensated Higher Order Ambisonics (NFC-HOA). For NFC-HOA no localization results are available. A few results are published for the similar method of Higher Order Ambisonics (HOA) which assumes plane waves as a physical model for the loudspeakers instead of point sources. For HOA, experiments were carried out for a central listening position,[10] and in some cases included off-center listening positions.[11,12] The results show a high dependency on the listening position. The best-achieved localization accuracy is 3° at a central listening position for fifth-order HOA employing 12 loudspeakers and a distance between the loudspeakers of approximately 2 m.

This paper investigates the localization accuracy for WFS and NFC-HOA, with a focus on the influence of

[a)]Portions of this work were published in H. Wierstorf, "Perceptual assessment of sound field synthesis," Ph.D. dissertation, Technische Universität Berlin, Berlin, Germany, 2014.
[b)]Electronic mail: hagen.wierstorf@tu-ilmenau.de
[c)]Present address: Audiovisual Technology Group, Technische Universität Ilmenau, Ilmenau DE-98693, Germany.

different loudspeaker spacings, different listening positions and different synthesized source types. It starts with a short review on the physical principles of WFS and NFC-HOA. Section II introduces the methodology of the experiment. This includes binaural simulations of the ear signals and the pointing method used to indicate the perceived direction. After that the paper presents and discusses the localization results for the different WFS and NFC-HOA configurations. The results are restricted to the horizontal plane as this is currently the most common sound presentation in sound field synthesis.

A. Sound field synthesis

The theory of sound field synthesis deals with the problem of finding driving signals for a loudspeaker array in order to achieve a desired sound pressure in a defined listening area. Mathematically this problem can be described by the single-layer potential.[13] It can be solved analytically for special loudspeaker array geometries. For circular or spherical geometries, the solutions are known as NFC-HOA and consist of a series expansion using basis functions up to an order of M. If M is restricted to half the number of applied loudspeakers, it is often called band-limited NFC-HOA, referring to a bandwidth limitation in the spatial domain. For arbitrary geometries the solution of the single-layer potential can be approximated for high frequencies by composing the loudspeaker array geometry from small planar surfaces for which analytical solutions can be found,[14] a method known as WFS. Because of the applied approximations only those parts of the loudspeaker array are active which emit sound in the propagation direction of the synthesized sound field.[15] There also exist numerical solutions to the single-layer potential.[16] Those are not considered in this paper as they in most cases represent solutions for well defined scenarios that cannot easily be generalized in contrast to the analytical solutions investigated in this study.

Before solving the single-layer potential, the desired sound field has to be defined. In this study, it is given by three different physical source models, point sources, focused sources, or plane waves. A focused source is a point source placed inside the listening area. This is accomplished by emitting a sound field by a subset of loudspeakers that travels toward a focus point and emanate afterward.[17] This implies that a focused source has a directivity and reaches its desired sound field only for listener positions that are not placed between active loudspeakers and the position of the focused source. The applied driving signals are listed in the Appendix for reference.

To demonstrate some of the properties of WFS and NFC-HOA, assume the synthesis of a mono-frequency plane wave with a frequency of 2 kHz using a circular loudspeaker array with a radius of 1.5 m, employing 56 loudspeakers with a distance of 17 cm between them. Figure 1 shows the resulting sound pressure using WFS, NFC-HOA with a high order, and spatial band-limited NFC-HOA with M chosen as half the number of loudspeakers. The synthesized sound field in the case of WFS and NFC-HOA with $M = 112$ is nearly identical, only the number of active loudspeakers differ between the two setups. In both cases, the sound field of the plane wave shows spatial aliasing artifacts in the upper part of the listening area. The artifacts arise due to the spatial under-sampling by the given number of loudspeakers which limits the sound field that can be synthesized correctly for higher frequencies. The frequency above which aliasing becomes prominent (aliasing frequency) can be approximated as $c/(2\Delta x_0)$, where Δx_0 is the loudspeaker spacing.[18] Because of the spatio-temporal nature of the problem the aliasing frequency is not only dependent on the loudspeaker spacing, but on others factors like the listening position. For most cases it is slightly higher for positions farther away from the most active loudspeakers. For spatially band-limited NFC-HOA with $M = 28$ the situation is different. Here, an alias-free region in the center of the array is visible surrounded by an area showing aliasing artifacts close to all loudspeakers. The size of that area is directly related to the distance between the loudspeakers.

The aliasing frequency and the corresponding artifacts are of special interest for the investigation of localization in sound field synthesis. The aliasing artifacts impair the interaural time differences (ITDs) and interaural level differences (ILDs) which are the main localization cues in the horizontal plane.[19] Nonetheless, if the aliasing frequency is above 1.4 kHz it can be hypothesized that the impaired cues for higher frequencies have no influence. This is based on the findings that the localization is dominated by lower frequency ITDs for broad band signals with static[20] or dynamic[21] cues. A corresponding lower bound of the aliasing frequency of 1.4 kHz would assume a loudspeaker spacing of 12 cm. This is in agreement with the listening test results for a central listening position.[3,6,7] As the aliasing

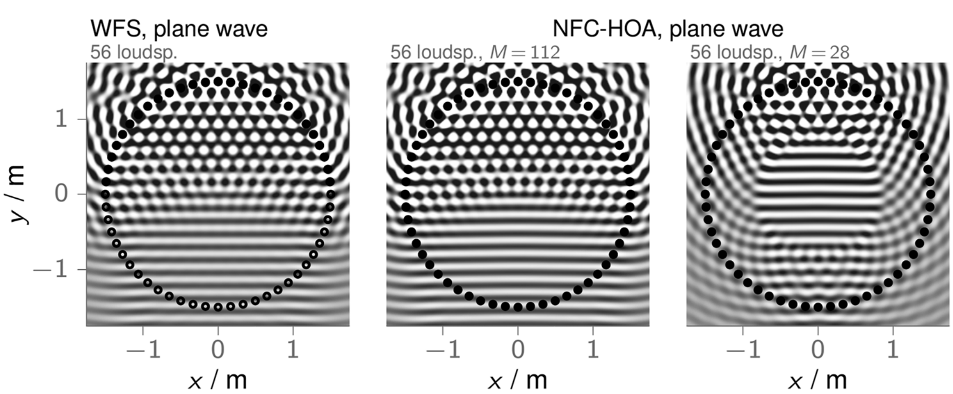

FIG. 1. Sound pressure of a plane wave with a frequency of 2 kHz traveling downward, synthesized by WFS and NFC-HOA with different orders M. The black dots indicate active loudspeakers, open circles inactive loudspeakers. The sound pressure is normalized at the center of the loudspeaker array, higher values are clipped.

artifacts change with listener position and source type, this study investigates if the localisation accuracy is similar in the whole listening area. Especially in the case of band-limited NFC-HOA, the deviations can be expected due to the structure of the sound field.

B. Reproducible research

The investigation and evaluation of sound field synthesis implies implementing a multitude of algorithms and running numerical simulations. As a consequence, the outcome of the algorithms is vulnerable to implementation errors which cannot completely be avoided.[22]

Beside the software tools, the work presented here relies on measured acoustical data. To ensure that other researchers can test the correctness of results and easily reproduce them,[23] the stimuli,[24] the results of the single listeners,[25] and the code for every single figure[26] are available as separate electronic publications. The involved signal processing based on the Sound Field Synthesis Toolbox,[27] which is a general framework for numerical simulation of WFS and NFC-HOA developed by the authors.

II. METHODS

The different loudspeaker arrays were simulated via dynamic binaural synthesis.[28] This process involved the presentation of the ear signals via headphones to the listeners while the signals were adapted according to the current head orientations. This allowed for a fast switching method between listening positions and avoided any hints to the listeners about their current position relative to the loudspeaker array. In the localization test itself, the listeners indicated the direction from which they perceived the synthesized sound by orienting their head to the perceived direction and pressing a key. A laser pointer was mounted on the headphones and provided them visual feedback about their head orientation. A pre-study, including ten participants, applied the exact same pointing method and dynamic binaural synthesis method of this study to the localization of 11 real loudspeakers placed behind an acoustic transparent curtain and of 11 binaurally rendered virtual sources placed at the same locations.[29] The result was an average localization accuracy of $1.8^\circ \pm 1.03^\circ$ for the real loudspeakers and $0.7^\circ \pm 1.13^\circ$ for the simulated once as long as the loudspeakers were positioned in a range of $\pm 35^\circ$. The difference between the averages of the perceived directions of the real and simulated loudspeakers was $1.6^\circ \pm 0.45^\circ$. This shows that both the accuracy of the pointing method and the error introduced by the dynamic binaural synthesis are similar to the human discrimination threshold and both can be applied to the investigation of the localization accuracy in sound field synthesis.

A. Dynamic binaural synthesis and stimuli

Stimuli were digitally generated at a sampling rate of 44.1 kHz. The impulse responses for the different loudspeaker setups and sound field synthesis conditions were calculated from a measured set of head-related transfer functions (HRTFs), which has a resolution of 1° and was measured with the KEMAR dummy head.[30] For non-measured directions, the HRTFs were linearly interpolated by a weighted sum of the two nearest measured HRTFs. For distances smaller or larger than the measured 3 m the HRTFs were adapted by delaying accordingly to the speed of sound and weighting inversely proportional to its distance. The changes that the different sound field synthesis methods would apply to every loudspeaker signal in order to synthesize a given source type were then included in the HRTFs. The SoundScape Renderer (SSR)[31] convolved the resulting HRTFs[24] with the audio material. As audio material, a train of Gaussian white noise pulses with a duration of 700 ms and a pause of 300 ms between each pulse was applied. The single pulses were independent white noise signals. They were windowed with raised cosine ramps of 20 ms length at their start and end. The resulting signal was band-pass filtered with a fourth order Butterworth filter between 125 and 20 000 Hz. It had a total length of 100 s and was stored and played back in a loop during the experiment. Figure 2 summarizes the dynamic binaural synthesis process.

The PC was equipped with a RME HDSP MADI card and for the digital to analog conversion CreamWare A16 converters were used. The listeners wore AKG K601 headphones, for which the HRTFs were compensated for by a non-individual filter designed after a KEMAR measurement and applying deconvolution and regularization.[32] The head movements of the listeners were tracked by a Fastrak Polhemus head tracker and passed onto the SoundScape Renderer with an update rate of 60 Hz. The SoundScape Renderer switched the HRTFs for the dynamic binaural synthesis, according to the orientation of the listener. This was performed on an audio block length of 1024 samples, resulting in an estimated latency of the whole dynamic binaural

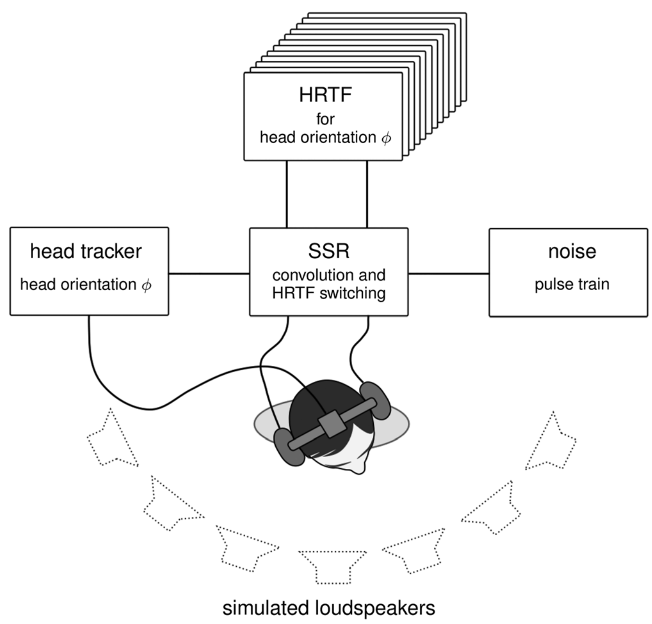

FIG. 2. Functional principle of dynamic binaural synthesis. The listener is wearing headphones and a head tracker. The audio material is convolved with the HRTF that incorporates all simulated loudspeakers for the corresponding listener orientation.

synthesis of around 120 ms, which was short enough to not affect the localization results.[33]

B. Pointing method

This paper applies a pointing method that Makous and Middlebrooks[34] introduced. Here, the listeners have to point with their heads toward the direction of the auditory event, while the sound event is present. This has the advantage that the listener is directly facing the source, a region in which the minimum audible angle is the smallest.[5] If the listeners are pointing their heads in the direction of the source during a closed-loop listening task without getting visual feedback about their looking direction, an estimation error of the sources at the side occurs, due to an interaction with the motor system.[35] To overcome this, a visual pointer was mounted onto the headphones in order to indicate the head orientation to the listener.

C. Procedure

The experiment utilized of three different circular loudspeaker setups. The radius of all loudspeaker arrays was 1.5 m. The center of the loudspeaker array was placed at (0, 0) m. The number of loudspeakers varied between 14, 28, and 56 loudspeakers. This corresponded to a loudspeaker spacing of 67, 34, 17 cm. For every loudspeaker setup, three different sound sources were synthesized by WFS using Eq. (A3), Eq. (A5), and Eq. (A7): a point source placed at $(0, 2.5)$ m, a plane wave traveling in the direction $(0, -1)$, and a focused source placed at $(0, 0.5)$ m. In addition, the same point source and plane wave were synthesized with NFC-HOA using Eq. (A1) and Eq. (A2). For the loudspeaker array consisting of 14 loudspeakers, two different Ambisonics orders $M = 7$ and $M = 28$ were applied. The test participants were placed at 16 different listening positions, which had a spacing of 25 cm along the x axis and 75 cm along the y axis. Because of the symmetry of the circular arrays, only listening positions on one half were considered. The overall setup is further highlighted by Fig. 3.

Three different loudspeaker setups, 16 different listening positions, and five different combinations of source type and sound field synthesis method, plus two different Ambisonics orders for one loudspeaker array resulted in a total of 288 conditions, that were presented five times to every listener. The experiment was split in four runs on different days. One run presented only WFS or NFC-HOA conditions in randomized order and lasted approximately 45 min.

The listeners sat on a chair in an acoustically damped listening room and had an acoustic transparent curtain 1.5 m in front of them. The room has a volume of 83 m^3 and a reverberation time RT_{60} of 0.17 s at a frequency of 1 kHz and was darkened during the experiment. The listeners wore headphones for the binaural presentation of the stimuli. They had a keyboard on their knees and a laser pointer was attached on top of the headphones. They were instructed to point in the direction from where they perceived the auditory event. The test participants were informed that the vertical direction should be ignored as the usage of non-individual HRTFs can lead to slightly elevated sources. After they made sure to point in the right direction, they were asked to hit the enter key. The stimuli were presented in a closed-loop to the listeners ending with the key press of the listeners, whereby there was no time limit for the response. The listeners' head orientation was calculated as the mean over the following ten values obtained from the head tracker, which corresponds to a time of 90 ms. After the key press, the next trial started instantaneously, which implied that the listener always started the localization from their last viewing direction. For the two NFC-HOA runs they were further instructed to look in the direction of the more pronounced source, if they heard more than one. In cases for which they were not able to state which was more pronounced, they were instructed that they should randomly choose one of the sources.

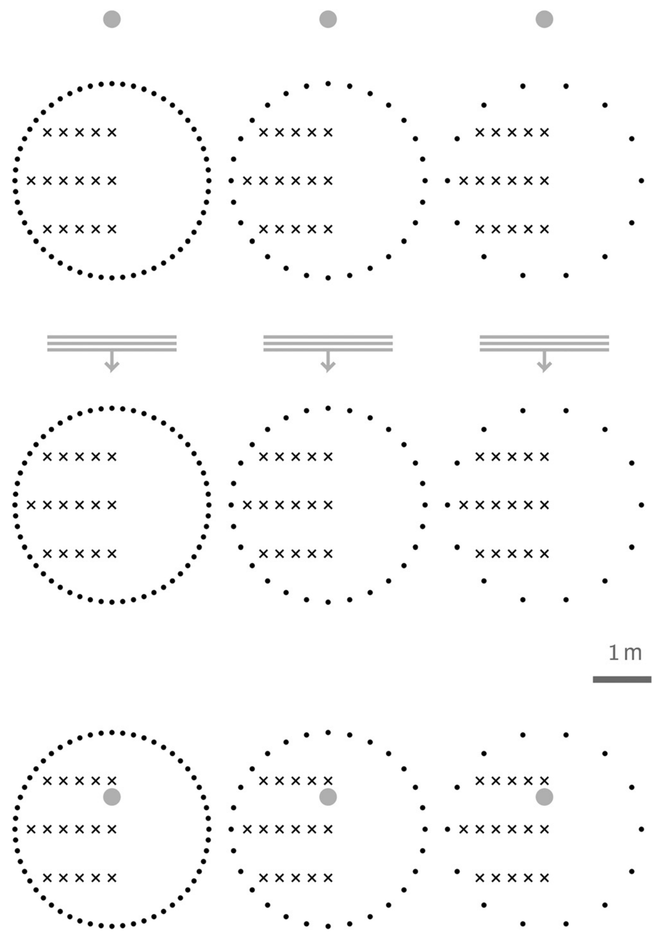

FIG. 3. Setup of the experiment. The position and type of the synthesized source is indicated by the grey symbols, the position of the listener by black crosses, and loudspeakers by black dots.

D. Listener

Twelve normal-hearing listeners were recruited for the two runs that included only WFS. They were aged 23 to 33 years old. One of them had prior experiences with psychoacoustic testing and sound field synthesis. Another 12 normal-hearing listeners were recruited for the two runs that included only NFC-HOA. They were aged 24 to 35 years old. Three of them had prior experiences with psychoacoustic testing and sound field synthesis. One of the listeners completed only the condition with plane wave as source model and one completed only the condition with point source as source model. Two test participants were excluded from the analysis, because their standard deviation of the reported direction over the five repetitions was more than twice as large as for the other participants.

III. RESULTS

Figure 4 summarizes the results[25] of all the experiments. For every sound field synthesis method, the loudspeaker positions are drawn as black dots and the synthesized sources are indicated by the grey symbols. At every listener position an arrow is pointing toward the average direction from which the listeners perceived the corresponding auditory event. The color of each arrow displays the localization error, which is defined as the absolute deviation between the desired sound event direction and the direction of the auditory event. It ranges from light for 0° to dark for values of 40° or higher.

The mean localization error for WFS synthesizing a point source or a plane wave is approximately 1° in the case of the loudspeaker spacing of 17 cm. Only at the position $(-1, 0.75)$ m for the synthesis of a plane wave the localization error is around 5°. For a loudspeaker spacing of 34 cm the localization error increases slightly to an average of 2°. For a loudspeaker spacing of 67 cm the localization error increases and varies for different listening positions, showing the largest errors at frontal listening positions. In addition, the listeners start to look in the direction of the nearest loudspeaker instead of the direction of the synthesized point source.

For the synthesis of a focused source in WFS, localization errors are in general larger. The focused source was placed at $(0, 0.5)$ m and was travelling downward, which means that the five frontal listener positions were placed between the focused source and the active loudspeakers, an area which should be avoided for focused sources. For these positions, the listeners were not able to perceive the direction of the focused but those of the active loudspeakers. In addition, it could be observed that for the loudspeaker arrays with a loudspeaker spacing of 67 and 34 cm only a small region with low localization error exists around the central listening positions. For positions to the side, the listeners were again pointing more in the direction of the active loudspeakers. Only for the loudspeaker array with 17 cm spacing between the loudspeakers, a triangle-shaped listening area can be identified where the localization error is around or less than 10°.

The localization error for band-limited NFC-HOA synthesizing a point source is larger at all listening positions than for WFS, in average 3.8° for a loudspeaker spacing of 17 cm and 7.4° for a spacing of 34 cm. The results are more dependent on the listening position as for the WFS conditions, showing stronger errors for positions to the side. In the case of the loudspeaker array with a loudspeaker spacing of 67 cm, the localization error for the point source condition is larger than 10° for most of the listening positions to the side. For the loudspeaker spacing of 67 cm, NFC-HOA with an order of $M = 28$ was also tested. In this case, the results are very similar to the ones of the WFS conditions for the same loudspeaker array. The localization error now has similar

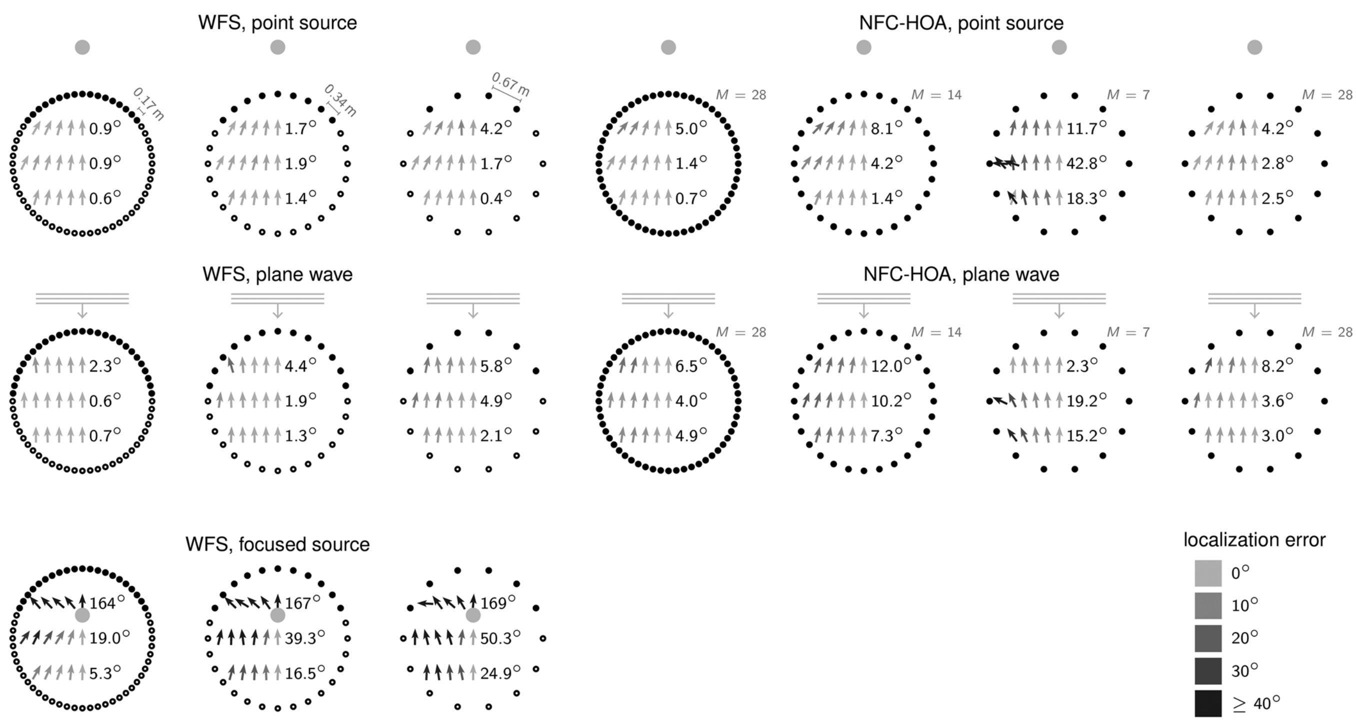

FIG. 4. Average localization results. The black circular symbols indicate loudspeakers, the grey ones the synthesized source. At every listening position, an arrow is pointing in the direction from which the listener perceived the corresponding auditory event. The color of the arrow displays the absolute localization error, which is also summarized as an average beside the arrows for every row of positions. The average confidence interval for all localization results is 2.3°.

values for all positions across the listening area and a small increase at frontal listening positions.

As most listeners reported that they sometimes perceived more than one source, the distributions of the reported directions of auditory events from all listeners were analyzed. First the mean standard deviation was calculated independently for each combination of synthesis method and source type. If a distribution for a given position had a standard deviation more than three times as large as the average it was identified as a condition giving an impression of more than one source. An example is presented in Fig. 5 for the point source condition at the listening position $(-1, -0.75)$ m and a loudspeaker spacing of 67 cm. The distributions of responds of nine listeners are shown in comparison for WFS, NFC-HOA with an order of $M = 28$, and band-limited NFC-HOA with an order of $M = 7$. For the case of WFS and NFC-HOA with an order of $M = 28$, a normal distribution with a standard deviation below 5° is visible. On the other hand, for band-limited NFC-HOA the distribution is far more spread with a standard deviation greater than the threshold of 19.3° and the data are most probably characterized by more than one normal distribution. In all cases where the standard deviation exceeded the threshold an expectation-maximization Gaussian mixture model was applied to estimate what data point belongs to what distribution. The two distributions and their corresponding data points are indicated by two different colors in Fig. 5. After the assignment to a particular distribution, the average direction of the auditory event was calculated for every distribution. In Fig. 4 two arrows, one for each corresponding direction were drawn for the three positions exceeding the standard deviation threshold. This was the case only for a point source synthesized by band-limited NFC-HOA with an order of $M = 7$ listening positions at the side.

IV. DISCUSSION

The accuracy of localizing a point source or plane wave that is synthesized by WFS is high in the whole listening area for all tested loudspeaker arrays. The localization error is below 5° on average and is only degraded for positions in the proximity to the active loudspeakers, where the aliasing frequency is slightly higher. This is in accordance with the results from Wierstorf *et al.*[9] There, high localization accuracy was achieved inside the listening area for a loudspeaker spacing of 20 cm. Only for loudspeaker spacings of more than 50 cm the localization was mainly bound to single loudspeakers comparable to the collapse of a stereophonic image into the closer loudspeaker. For a focused source, the localization accuracy depends more on the actual listening positions. Only in a very small part of the listening area is the localization accuracy below 5°.

The localization accuracy for the same loudspeaker setups driven by band-limited NFC-HOA is inferior to that of WFS. Only the loudspeaker array with a loudspeaker spacing of 17 cm is capable of providing a localization error smaller than 5° in most of the listening area. For fewer loudspeakers, large localization errors occur outside the center of the listening area and listeners can perceive more than one source and hear single loudspeakers. If the order of NFC-HOA is increased, the localization accuracy is comparable to that of WFS. This highlights that for NFC-HOA the applied order is very critical for the localization accuracy in the whole listening area. If the order is reasonably high, localization is identical to that of WFS. Otherwise it is better inside and worse outside of the center of the listening area.

The described results can be further discussed by estimating relevant binaural cues by a binaural model.[36] ITD, ILD, and interaural coherence (IC) values were extracted from the binaural signals for each auditory filter using rectangular time windows of 20 ms length. ITD and IC values were calculated from the cross-correlation function. ITD- and ILD-histograms were accumulated over the time of 700 ms of one noise pulse. The histogram bin sizes were 50 μs and 1 dB, respectively. An ITD/ILD sample belonging to a specific bin was weighted by its corresponding IC (between 0 and 1) value in order to incorporate reliability of that sample.[37]

Figure 6 presents the ITD and ILD histograms for a listener placed at $(-1, -0.75)$ m for the different sound field synthesis systems and different source types. Up to 1.3 kHz only ITDs were considered and for larger frequencies only ILDs. Most of the WFS conditions were able to resemble the ITD pattern of the reference source, which seemed also to dominate the perceived direction of the listeners. This is further highlighted by another study of the authors. There, they used ITD values below 1.4 kHz to predict the perceived directions of synthesized point sources in WFS and a linear loudspeaker array with high accuracy.[9] For the conditions of WFS synthesizing a focused source, the situation was different and two effects were visible. For low frequencies the ITDs were less reliable and spread around a larger range

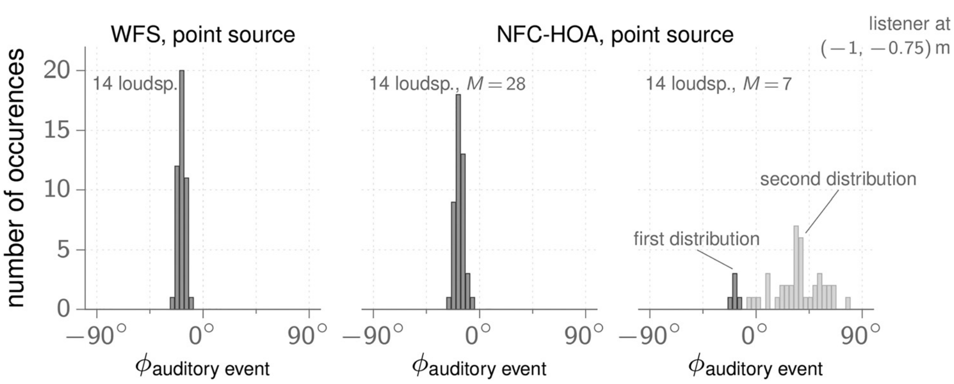

FIG. 5. Example distributions of the directions of the auditory event. The directions were judged by nine listeners at the position $(-1, -0.75)$ m for a loudspeaker array with a loudspeaker spacing of 67 cm. The results for a synthesized point source for WFS and NFC-HOA for different orders M are shown.

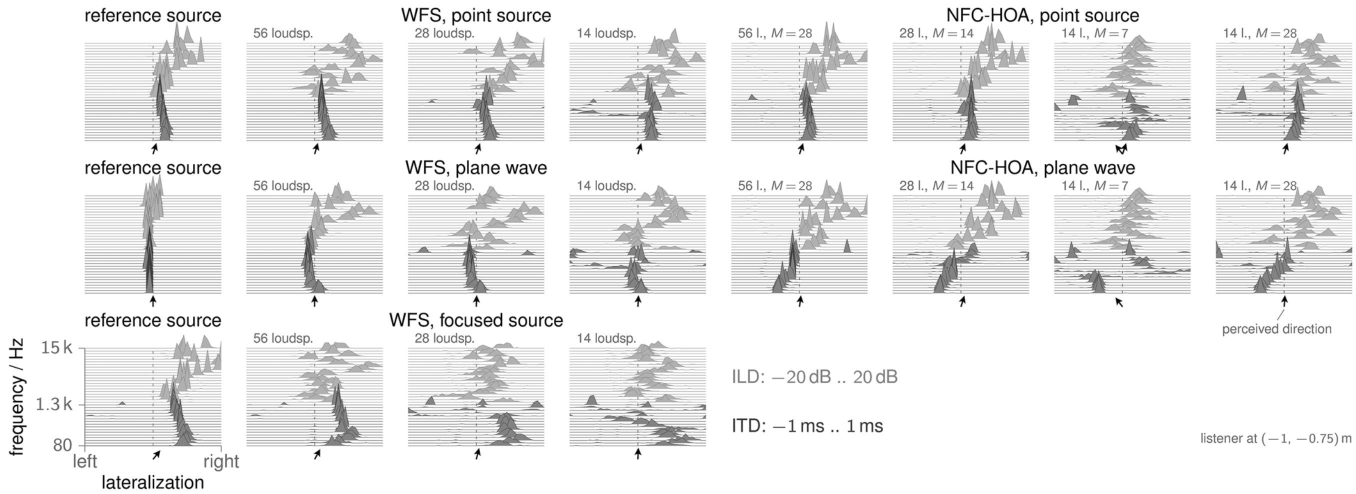

FIG. 6. ITD and ILD histograms over one noise pulse for different center frequencies. The ITDs were calculated up to 1.3 kHz, ILDs above. The results are shown for the listener position $(-1, -0.75)$ m inside different sound fields. Those sound fields consisted of a reference case and the ones where it was tried to recreate the reference sound field with WFS and NFC-HOA. The black arrow underneath each graph points in the direction of the auditory event reported by the listeners. For the reference cases the direction of the sound event is used.

than at higher frequencies. This is due to the limitations in acoustic focusing at lower frequencies.[38] The second effect showed a spread of ITD values for higher frequencies, accompanied by smaller ILD values. This was more pronounced for loudspeaker arrays with larger spacings. Here, the ITDs for high frequencies were more influenced by the positions of the nearest active loudspeakers than by the position of the focused source. This is visible in the responds of the listeners in Fig. 4 as well, leading to a localization of the active loudspeakers for loudspeaker arrays with large spacings and listening positions to the side.

ITDs of the signals synthesized by NFC-HOA showed slightly larger deviations from the ITDs of the reference signals than in WFS. Interestingly, the amount of deviations were not only related to the distance between the loudspeakers, but to the applied order as well. Both sound fields synthesized with band-limited NFC-HOA and an order of $M = 7$ showed a wide spread of ITD values that corresponded to the perception of two sources in the case of the synthesized point source as shown in Fig. 5. In addition, for the synthesis of a plane wave there was in general more spread between the ITD of the lowest and the highest analyzed frequency band than for a point source. In general, binaural cues available in the synthesized sound fields were less reliable the larger the distances between the loudspeakers and the lower the Ambisonics order.

V. CONCLUSION

Sound field synthesis methods like WFS and NFC-HOA target at physically controlling a sound field in an extended listening area surrounded by loudspeakers. Because of the limited number of loudspeakers used in common setups, the sound fields show artifacts above a given frequency. This study investigated how those artifacts influence the localization accuracy of a listener. The results show that even with relatively low numbers of loudspeakers and loudspeaker spacings around 20 cm a localization accuracy could be reached in the whole listening area that is comparable to the one achieved in natural sound fields. For larger loudspeaker spacings of around 70 cm the localization accuracy decreases to 5°. For band-limited NFC-HOA the localization becomes position dependent with higher accuracy in the center of the listening area and worse localization accuracy, even splitting of sources, outside of the center.

Most findings can be well explained by the modifications the sound field artifacts introduce to the low-frequency ITDs at the listener ears. This was further highlighted by Wierstorf *et al.*[9] who used a binaural model to predict localization results for a linear loudspeaker array based on ITDs.

This work lays a foundation for the investigation of the larger question of how sound field synthesis systems influence the sound quality experienced by the listener. To tackle this question future work should include the influence of the aliasing artifacts on the perceived timbre as well.[4,39]

ACKNOWLEDGMENTS

The authors wish to thank Matthias Geier for providing help with implementations for the experiment, and Fiete Winter for discussions on binaural modelling. We also thank two anonymous reviewers for their very helpful and supportive comments. This work was supported by EU FET Grant Two!Ears No. ICT-618075.

APPENDIX: DRIVING SIGNALS

Because of the usage of a circular loudspeaker array the synthesis is called 2.5-dimensional.[3] The loudspeakers show the characteristics of 3-dimensional point sources. A circular loudspeaker array on the other hand states a 2-dimensional problem. This implies that the amplitude is not correct in the whole listening area, because energy is floating out of the 2-dimensional area. This is handled for 2.5-dimensional driving signals by adding a reference point $\mathbf{x}_{\mathrm{ref}}$ at which the desired amplitude is ensured. In the following the driving signals used in this study are listed, the interested reader is referred to Wierstorf[40] for a detailed derivation of them.

For a circular loudspeaker array with radius r_0, $\mathbf{x}_{\text{ref}} = (0,0)$ and a point source as source model the 2.5-dimensional NFC-HOA driving function is given as

$$D_{\text{NFC-HOA,ps}}(\phi_0,\omega) = A(\omega)\frac{1}{2\pi r_0}\sum_{m=-M}^{M}\frac{h_{|m|}^{(2)}\left(\frac{\omega}{c}r_s\right)\Phi_{-m}(\phi_s)}{h_{|m|}^{(2)}\left(\frac{\omega}{c}r_0\right)}\Phi_m(\phi_0), \quad \text{(A1)}$$

where ω denotes the angular frequency, (ϕ_0, r_0) is the position of the loudspeaker, (ϕ_s, r_s) the position of the point source, $h_{|m|}^{(2)}$ the spherical Hankel function of second kind and order $|m|$, $\Phi_m(\phi) = e^{im\phi}$ a circular basis function, $A(\omega)$ the amplitude spectrum of the source model, c the speed of sound, and M the Ambisonics order.

For a circular loudspeaker array with radius r_0, $\mathbf{x}_{\text{ref}} = (0,0)$ and a plane wave as source model the 2.5-dimensional NFC-HOA driving function is given as

$$D_{\text{NFC-HOA,pw}}(\phi_0,\omega) = A(\omega)\frac{2i}{r_0}\sum_{m=-M}^{M}\frac{i^{-|m|}\Phi_{-m}(\phi_k)}{\frac{\omega}{c}h_{|m|}^{(2)}\left(\frac{\omega}{c}r_0\right)}\Phi_m(\phi_0), \quad \text{(A2)}$$

where ϕ_k is the direction of the plane wave.

For a point source as source model the 2.5-dimensional WFS driving function is given as

$$D_{\text{WFS,ps}}(\mathbf{x}_0,\omega) = -\sqrt{\frac{\mathbf{x}_{\text{ref}}}{2\pi}}A(\omega)w(\mathbf{x}_0)\sqrt{i\frac{\omega}{c}} \times \frac{\langle \mathbf{x}_0 - \mathbf{x}_s, \mathbf{n}_{\mathbf{x}_0}\rangle}{|\mathbf{x}_0 - \mathbf{x}_s|^{\frac{3}{2}}}e^{-i(\omega/c)|\mathbf{x}_0-\mathbf{x}_s|}, \quad \text{(A3)}$$

where $\mathbf{x}_0$ is the position of the loudspeaker, $\mathbf{x}_s$ the position of the point source, $\mathbf{n}_{\mathbf{x}_0}$ a normal vector pointing from a single loudspeaker to the center of the array, and $w(\mathbf{x}_0)$ the loudspeaker selection function with

$$w(\mathbf{x}_0) = \begin{cases} 1, & \langle \mathbf{x}_0 - \mathbf{x}_s, \mathbf{n}_{\mathbf{x}_0}\rangle > 0 \\ 0, & \text{else.} \end{cases} \quad \text{(A4)}$$

For a plane wave as source model the 2.5-dimensional WFS driving function is given as

$$D_{\text{WFS,pw}}(\mathbf{x}_0,\omega) = -\sqrt{8\pi\mathbf{x}_{\text{ref}}}\,w(\mathbf{x}_0)A(\omega) \times \sqrt{i\frac{\omega}{c}}\langle \mathbf{n}_k, \mathbf{n}_{\mathbf{x}_0}\rangle e^{-i(\omega/c)\mathbf{n}_k\mathbf{x}_0}, \quad \text{(A5)}$$

where $\mathbf{n}_k$ denotes the direction of the plane wave and the loudspeaker selection is given by

$$w(\mathbf{x}_0) = \begin{cases} 1, & \langle \mathbf{n}_k, \mathbf{n}_{\mathbf{x}_0}\rangle > 0 \\ 0, & \text{else.} \end{cases} \quad \text{(A6)}$$

For a focused source as source model the 2.5-dimensional WFS driving function is given as

$$D_{\text{WFS,fs}}(\mathbf{x}_0,\omega) = -\sqrt{\frac{\mathbf{x}_{\text{ref}}}{2\pi}}A(\omega)w(\mathbf{x}_0)\sqrt{i\frac{\omega}{c}} \times \frac{\langle \mathbf{x}_0 - \mathbf{x}_s \mathbf{n}_{\mathbf{x}_0}\rangle}{|\mathbf{x}_0 - \mathbf{x}_s|^{\frac{3}{2}}}e^{i(\omega/c)|\mathbf{x}_0-\mathbf{x}_s|}. \quad \text{(A7)}$$

The loudspeaker selection is given by including the traveling direction $\mathbf{n}_s$ of the focused source as

$$w(\mathbf{x}_0) = \begin{cases} 1, & \langle \mathbf{n}_s, \mathbf{x}_s - \mathbf{x}_0\rangle > 0 \\ 0, & \text{else.} \end{cases} \quad \text{(A8)}$$

[1] J. C. Steinberg and W. B. Snow, "Symposium on wire transmission of symphonic music and its reproduction in auditory perspective: Physical factors," Bell Syst. Tech. J. **13**, 245–258 (1934).

[2] F. M. Fazi and P. A. Nelson, "Sound field reproduction as an equivalent acoustical scattering problem," J. Acoust. Soc. Am. **134**, 3721–3729 (2013).

[3] E. Start, "Direct sound enhancement by wave field synthesis," Ph.D. dissertation, Technische Universiteit Delft, Delft, the Netherlands, 1997.

[4] H. Wierstorf, C. Hohnerlein, S. Spors, and A. Raake, "Coloration in wave field synthesis," in *55th International Conference of the Audio Engineering Society*, Helsinki, Finland (2014), preprint 5-3.

[5] A. W. Mills, "On the minimum audible angle," J. Acoust. Soc. Am. **30**, 237–246 (1958).

[6] P. Vogel, "Application of wave field synthesis in room acoustics," Ph.D. dissertation, Technische Universiteit Delft, Delft, the Netherlands, 1993.

[7] H. Wittek, "Perceptual differences between wavefield synthesis and stereophony," Ph.D. dissertation, University of Surrey, Surrey, UK, 2007.

[8] E. Verheijen, "Sound reproduction by wave field synthesis," Ph.D. dissertation, Technische Universiteit Delft, Delft, the Netherlands, 1997.

[9] H. Wierstorf, A. Raake, and S. Spors, "Binaural assessment of multichannel reproduction," in *The Technology of Binaural Listening*, edited by J. Blauert (Springer, New York, 2013), Chap. 10, pp. 255–278.

[10] S. Bertet, J. Daniel, E. Parizet, and O. Warusfel, "Investigation on localisation accuracy for first and higher order ambisonics reproduced sound sources," Acta Acust. **99**, 642–657 (2013).

[11] M. Frank, F. Zotter, and A. Sontacchi, "Localization experiments using different 2D ambisonics decoders," in *25th Tonmeistertagung - VDT International Audio Convention*, Leipzig, Germany (2008), pp. 696–704.

[12] P. Stitt, S. Bertet, and M. Walstijn, "Off-centre localisation performance of Ambisonics and HOA for large and small loudspeaker array radii," Acta Acust. **100**, 937–944 (2014).

[13] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory* (Springer, New York, 1998).

[14] F. M. Fazi, P. A. Nelson, and R. Potthast, "Analogies and differences between three methods for sound field reproduction," in *1st Ambisonics Symposium*, Graz, Austria (2009).

[15] D. W. Herrin, F. Martinus, T. W. Wu, and A. F. Seybert, "A new look at the high frequency boundary element and rayleigh integral approximations," in *Noise & Vibration Conference and Exhibition*, Traverse City, MI (2003), paper 1451.

[16] O. Kirkeby and P. A. Nelson, "Reproduction of plane wave sound fields," J. Acoust. Soc. Am. **94**, 2992–3000 (1993).

[17] S. Yon, M. Tanter, and M. Fink, "Sound focusing in rooms: The time-reversal approach," J. Acoust. Soc. Am. **113**, 1533–1543 (2003).

[18] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," J. Acoust. Soc. Am. **93**, 2764–2778 (1993).

[19] J. Blauert, *Spatial Hearing* (The MIT Press, Cambridge, MA, 1997).

[20] F. L. Wightman and D. J. Kistler, "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661 (1992).

[21] E. A. Macpherson, "Cue weighting and vestibular mediation of temporal dynamics in sound localization via head rotation," Proc. Mtgs. Acoust. **19**, 050131 (2013).

[22]D. C. Ince, L. Hatton, and J. Graham-Cumming, "The case for open computer programs," Nature **482**, 485–488 (2012).
[23]D. L. Donoho, A. Maleki, I. U. Rahman, M. Shahram, and V. Stodden, "Reproducible research in computational harmonic analysis," Comput. Sci. Eng. **11**, 8–18 (2009).
[24]H. Wierstorf, "Binaural room scanning files for sound field synthesis localization experiment," Data set, Zenodo (2016), doi:10.5281/zenodo.55427.
[25]H. Wierstorf, "Listening test results for sound field synthesis localization experiment," Data set, Zenodo (2016), doi: 10.5281/zenodo.55439.
[26]H. Wierstorf, "Code to reproduce the figures in the paper 'Assessing localization accuracy in sound field synthesis,' " Software, Zenodo (2016), doi:10.5281/zenodo.166755.
[27]H. Wierstorf and S. Spors, "Sound field synthesis toolbox," in *132nd Convention of the Audio Engineering Society*, Budapest, Hungary, 2012, No. 50, http://matlab.sfstoolbox.org (Last viewed 2/16/2017).
[28]U. Horbach, A. Karamustafaoglu, R. Pellegrini, P. Mackensen, and G. Theile, "Design and Applications of a Data-based Auralization System for Surround Sound," in *106th Convention of the Audio Engineering Society*, Munich, Germany, 1999, preprint 4976.
[29]H. Wierstorf, S. Spors, and A. Raake, "Perception and evaluation of sound fields," in *59th Open Seminar on Acoustics*, Boszkowo, Poland, 2012, pp. 263–268.
[30]H. Wierstorf, M. Geier, A. Raake, and S. Spors, "A free database of head-related impulse response measurements in the horizontal plane with multiple distances," in *130th Convention of the Audio Engineering Society*, London, UK, 2011, No. 6.
[31]M. Geier, J. Ahrens, and S. Spors, "The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods," in *124th Convention of the Audio Engineering Society*, Amsterdam, the Netherlands, 2008, preprint 7330.
[32]O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," IEEE Trans. Speech Audio Proc. **6**, 189–195 (1998).
[33]E. M. Wenzel, "Effect of increasing system latency on localization of virtual sounds," in *16th International Conference of the Audio Engineering Society*, Rovaniemi, Finland, 1999, preprint 16-004.
[34]J. C. Makous and J. C. Middlebrooks, "Two-dimensional sound localization by human listeners," J. Acoust. Soc. Am. **87**, 2188–2200 (1990).
[35]J. Lewald, G. J. Dörrscheidt, and W. H. Ehrenstein, "Sound localization with eccentric head position," Behav. Brain. Res. **108**, 105–125 (2000).
[36]R. Decorsière, T. May, C. Kim, and H. Wierstorf, "Two!Ears Auditory Front-end 1.0," Softwarte, Zenodo (2015), doi:10.5281/zenodo.28008.
[37]C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," J. Acoust. Soc. Am. **116**, 3075–3089 (2004).
[38]R. Oldfield, "The analysis and improvement of focused source reproduction with wave field synthesis," Ph.D. dissertation, University of Salford, Salford, UK, 2013, Secs. 4.7 and 4.9.
[39]F. Rumsey, S. Zielinski, R. Kassier, and S. Bech, "On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality," J. Acoust. Soc. Am. **118**, 968–976 (2005).
[40]H. Wierstorf, "Perceptual assessment of sound field synthesis," Ph.D. dissertation, Technische Universität Berlin, Berlin, Germany, 2014.